Data Paper

CanFlyet: habitat zone and diet trait dataset for Diptera species of Canada and Greenland

Samantha E Majoros[‡], Tyler A Elliott[‡], Sarah J. Adamowicz[§]

‡ University of Guelph, Guelph, Canada
§ Biodiversity Institute of Ontario, Guelph, Guelph, Canada

Abstract

Background

Flies (Diptera) are an ecologically important group that play a role in agriculture, public health and ecosystem functioning. As researchers continue to investigate this order, it is beneficial to link the growing occurrence data to biological traits. However, large-scale ecological trait data are not readily available for fly species. While some databases and datasets include fly data, many ecologically relevant traits for taxa of interest are not included. In this study, we create a dataset containing ecological traits (habitat and diet) for fly species of Canada and Greenland having occurrence records on the Barcode of Life Data Systems (BOLD). We present a dataset containing trait information from the literature for 981 Diptera species.

New information

Diptera species were chosen for the dataset, based on the occurrence records available for Diptera species from Canada and Greenland on the Barcode of Life Data System (BOLD). Trait data were then compiled and digitised in a standardised format, based on

667 works from literature published before April 2024. Traits were assigned at the lowest taxonomic level available. Three biological traits were included: larval habitat, larval diet type and adult diet. The dataset contains traits for 981 species across 380 genera, 34 subfamilies and 61 families. This dataset allows for assignment of traits to occurrence data for Diptera species and can be used for further research into the ecology, evolution and conservation of this order.

Keywords

Diptera, Canada, Greenland, dataset, habitat, diet, ecological traits

Introduction

Flies (Diptera) are a diverse, widespread taxonomic group, occurring across a wide range of ecological niches and geographic regions (Marshall 2012,Adler and Courtney 2019). This important group occupies various ecological roles, many of which have an impact on the environment and agricultural practices. Fly species can act as pests to crops, such as *Liriomyza bryoniae* on cabbages, lettuce and tomatoes (Bragard et al. 2020); *Delia radicum* on broccoli and cauliflower (Mesmin et al. 2019); *Dasineura mali* on apples (Wearing et al. 2013); and *Bradysia ocellaris* and *Lycoriella ingenua* on mushrooms (Shamshad 2010). Diptera also includes pests of livestock, such as the blood-feeding *Stomoxys calcitrans* (Olafson et al. 2021). Flies are also a group of interest regarding public health, as some, like *Calliphora vomitoria*, are vectors of pathogenic microorganisms (Bedini et al. 2017). Despite these detrimental examples, flies also provide many important ecological benefits. Flies are an important pollinator group, especially in Arctic and northern environments (Tiusanen et al. 2016). The family Muscidae includes species that are pollinators of key plant species in these regions and play an important role in the ecosystem (Tiusanen et al. 2016). Flies also fill various other ecological roles, such as parasitoids of other insects that act as vectors for pathogenic bacteria (Molinatto et al. 2020). Many species act as ecosystem engineers; changing environments through suspension-feeding, grazing, burrowing and predation, as well as serving as an important food source for other organisms (Adler and Courtney 2019). The saprophagous diet of many flies is also important for the breakdown of decaying organic material (Adler and Courtney 2019). As biocontrol agents and bioindicators of water quality, flies act as a valuable resource for food production and research (Adler and Courtney 2019).

This ecologically important group has been the focus of large DNA barcoding efforts, such as the Global Malaise Trap Program, which seeks to document Diptera diversity (which is likely under-described) and open doors for ecological research and monitoring (Geiger et al. 2016, deWaard et al. 2018, Global Malaise Trap Program 2023). DNA barcoding involves using a standardised gene region to identify and delineate species (Hebert et al. 2003) and there is a large amount of DNA barcoding data available on databases such as the Barcode of Life Data System (BOLD) (Ratnasingham and Hebert

2007), which contained 6,660,909 Diptera records representing 36,125 species as of 2 Oct 2024. There is also a large amount of species occurrence data on databases such as the Global Biodiversity Information Facility (GBIF) (GBIF.org 2024), which contained 24,618,171 occurrences as of 2 Oct 2024. As Diptera research continues, it is valuable to link such occurrence records with biological data.

However, finding biological trait data for fly species can be a challenge. There are various databases or datasets currently available for Diptera and other insect groups, covering a varying number of traits and taxa (Table 1). However, few databases focus solely on Diptera and, while flies are included in some databases, not all species and traits of interest are included. A larger focus also appears to be on morphological traits, such as body size, with information on ecological traits, such as habitat and diet, being less readily available. Finally, few databases focus on Diptera from Canada or Greenland. Overall, traits for many Diptera species are very difficult to find and, when known, are typically present in taxon-specific studies and prose format rather than compiled datasets with standardised digitisation.

Table 1.

Overview of the publicly available databases and datasets that include insects. This table includes databases and datasets that were publicly available as of 2 Feb 2023 and were found through a literature search. Number of records are as of 2 Feb 2023.

Database Name	Taxonomic Focus	Geographic Focus	Types of Traits	Number of Traits	Number of Records	Data Availability	References
DISPERSE[*1]	Annelida, Mollusca, Platyhelminthes & Arthropoda	Europe	Dispersal	9	480	Download as xlsx file	Sarremejane et al. (2020)
Freshwater Biological Traits[*2]	Freshwater macroinvertebrate taxa	North America	Morphology, life history, mobility, environmentaltolerance, resource acquisition/ preference	26	11912	Select taxa, region and traits on website. Download as an xlsx file	United States Environmental Protection Agency (2012)
Animal Traits[*3]	Arthropoda, Chordata, Mollusca and Annelida	Global	Body mass, metabolic rate, brain size	3	3580	Download xlsx or csv file	Herberstein et al. (2022)
CONUS[*4]	Arthropoda	United States of America	Life history, dispersal, morphology, ecology	11	2.05 million	Download csv file	Twardochleb et al. (2021)
Lotic Invertebrate Traits for North America[*5]	Invertebrates	North America	Ecology, morphology, behaviour, physiology	62	14127	Download as a txt file	Vieira et al. (2006)

Database Name	Taxonomic Focus	Geographic Focus	Types of Traits	Number of Traits	Number of Records	Data Availability	References
European & Maghreb Butterfly Trait Database[6]	Lepidoptera	Europe and North Africa	Life history, morphology, resource-based, behaviour	25	542	Download as xlsx file	Middleton-Welling et al. (2020)
GlobalAnts[7]	Formicidae	Global	Morphology, ecology, life history	23	82910	Need to make an account to access. Can search for desired traits and download as a csv file	Parr et al. (2016)
The Odonate Phenotypic Database[8]	Odonata	Global	Morphology, life history, behaviour	35	3978	Download as a xlsx file	Waller et al. (2019)
Morphological trait database of saproxylic beetles[9]	Saproxylic Coleoptera	Europe	Morphology	13	1376	Download as a csv file	Hagge et al. (2021)
Data from: Sensitivity of functional diversity metrics to sampling intensity[10]	Carabidae	The Netherlands	Morphology	7	73	Download as txt file	van der Plas et al. (2017)
Data from: A summary of eight traits of Coleoptera, Hemiptera, Orthoptera and Araneae, occurring in Grasslands in Germany[11]	Arthropoda	Germany	Morphology, ecology	8	1230	Download as txt file	Gossner et al. (2015)
LepTraits[12]	Lepidoptera	Global	Morphology, habitat, reproduction, hostplant association	6	75103	Download as csv file	Shirey et al. (2022)
The Insect Trait Tool (ITT)[13]	Arthropoda	Germany	Habitat, Diet	2	34085	Download as a pdf or .xlsx file	Hörren et al. (2022)

The goal of this study was to create a dataset containing ecological traits (habitat and diet type) for selected Diptera species found in Canada and Greenland. These traits were selected due to the paucity of such information in available datasets and the value of this

information for understanding shifts in traits and not only taxonomic composition, across space and time. Species were selected that have multiple occurrence records and high-quality DNA barcode sequences on BOLD (following criteria outlined in Majoros et al. (2023)), which likely includes many of the most common species in the study region. This trait dataset is expected to be of particular relevance for species caught in Malaise traps, given the extensive usage of this collection modality in large-scale DNA barcoding efforts. This dataset can be used in future research to further our knowledge and understanding of this important taxonomic group, including ecology and temporal studies of insect communities with climate and land-use changes.

General description

Purpose: This study provides a dataset containing the larval habitat and adult and larval diet categories for Diptera species of Canada and Greenland having occurrence records on BOLD. The traits were assigned to taxa using the currently available literature. This dataset will aid in further research that involves the ecological traits of Diptera species.

Project description

Title: CanFlyet: Habitat Zone and Diet Trait Dataset for Diptera Species of Canada and Greenland

Personnel: Samantha E. Majoros, Tyler A. Elliott and Sarah J. Adamowicz

Design description: The dataset provides larval habitat, larval diet and adult diet categories for Diptera species found in Canada and Greenland. Traits were determined based on a literature search.

Funding: Funding support for this project comes from the Natural Sciences and Engineering Research Council of Canada; the Government of Canada through Genome Canada and Ontario Genomics; the Ontario Ministry of Economic Development, Job Creation and Trade; and Food from Thought: Agricultural Systems for a Healthy Planet Initiative programme funded by the Government of Canada through the Canada First Research Excellence Fund.

Sampling methods

Description: The fly species were chosen for inclusion in this dataset by first downloading data for Diptera from Canada and Greenland from BOLD directly into R using BOLD's application programming interface (API) on 24 June 2021. This dataset was originally used in Majoros et al. (2023), in which species were represented by Barcode Index Numbers (BINS), which are operational taxonomic units (OTUs) used by BOLD that are clusters of barcode sequences similar to species (Ratnasingham and Hebert 2013). In addition to BINs, the analysis was also repeated by clustering records by

a 4% clustering threshold and using these clusters in the place of BINs. Studies have shown that a 3-5% clustering threshold is suitable for clustering within the Diptera family Chironomidae (Lin et al. 2015, Baloğlu et al. 2018). The records were filtered based on the requirements outlined in Majoros et al. (2023), which included that records needed to possess high-quality DNA sequence data and be identified to the species or genus level, that BINs and clusters needed to possess at least 20 records and be found in at least two geographic regions meeting the requirements of Majoros et al. (2023) and needed to be represented by at least 10 records in each region in which they were found. The remaining species represented by these BINs and clusters were chosen for analysis and inclusion in this dataset. The traits were assigned to taxa using a series of literature searches conducted between April 2021 and April 2024. The dataset was formatted using the DarwinCore format (Darwin Core Maintenance Group 2023).

Sampling description: The traits included in this dataset are larval habitat and adult and larval diet categories. These traits were chosen due to their importance in ecological functioning as well as the impact these traits can have on species distributions and population genetic structure. These traits were used in Majoros et al. (2023) as part of the case study. The biological traits for each species were determined and assigned through literature searches conducted from April 2021 to April 2024. Through the Omni Academic search tool available through the University of Guelph and Google Scholar, traits were found using the following search terms: trait AND "Taxonomic name", habitat AND "Taxonomic name", diet AND "Taxonomic name", "Feeding mode" AND "Taxonomic name", biology AND "Taxonomic name", "Natural history" AND "Taxonomic name", "Life history" AND "Taxonomic name", taxonomy AND "Taxonomic name", catalogue AND "Taxonomic name" and "Field guide" AND "Taxonomic name". This dataset compiles information from 667 published sources. The title and DOI of each reference are included in the dataset and the full references are available in Suppl. material 1. Traits were assigned to the lowest taxonomic level possible; however, not all traits could be assigned at the species level. For these species, traits were assigned using data from the next lowest level available, whether genus, subfamily or family. For example, if the authors of a given study mentioned that a particular genus has a specific trait, then species belonging to that genus were assigned that trait. The taxonomic level of the trait data is included in the dataset.

For larval habitat, species were classified as terrestrial (defined as taxa that live primarily in land habitats), aquatic (defined as taxa that live primarily in waterbodies or associated habitats) or semi-aquatic (defined as species that primarily live in wet habitats and require high levels of moisture and some elements of both terrestrial and aquatic habitats). These assignments were made based on the larval habitat requirements. For adult and larval diet, species could be classified as predaceous (taxa that prey on insects or other animals), mycophagous (those that feed on fungi), saprophagous (those that feed on decaying organic matter), nectar/pollen/honeydew feeding (taxa that primarily feed on nectar, pollen and/or honeydew), parasitic (those that live and feed in or on an organism of another species), parasitoid (organism whose young develop within a host, eventually resulting in host death), leaf/root/stem feeding (those that feed on the leaves,

roots and/or stems of plants), detritus and algae feeding (those that feed on small particles of algae and detritus), kleptoparasitic (those that take food and resources from another species), polyphagous (those that engage in more than three of the other feeding modes), non-feeding (those that do not feed at the specified life stage) and unclear (taxa for which there is not enough information to make a trait assignment). Taxa can also be assigned a combination of the diet categories. Many higher taxonomic levels are diverse and contain species with different traits. In this case, the most common category was used if specific information for a species was not available. When this was unclear, multiple categories were included separated by "or".

Quality control: The traits in the dataset were assigned on the basis of reading the literature by a single researcher. To check the trait assignments for accuracy and consistency, the dataset was sent to another researcher for review. This second researcher randomly selected 20 species from the dataset and reviewed the referenced literature within the dataset for the species to ensure that the trait assignments were accurate and consistent. Knowledge of fly traits is likely to develop as further research is conducted and we are open to suggestions from the community, via comments on the manuscript, dataset or GitHub and will update the dataset as needed.

Geographic coverage

Description: This dataset includes Diptera species found in Canada and Greenland that also have multiple occurrence records and DNA barcode sequences publicly available on BOLD. Ten of the species are unique to Greenland and the rest are found in Canada.

Taxonomic coverage

Description: The dataset contains traits for 981 species across 380 genera, 34 subfamilies and 61 families. The families are represented by varying numbers of species within the dataset (Fig. 1). For several families, the dataset only contains information for one species. The most species-rich family in this dataset is Chironomidae, which is represented by 182 species.

Traits coverage

The dataset contains information for three biological traits: larval habitat, adult diet and larval diet. For larval habitat, 655 species were considered terrestrial, 277 were aquatic and 58 were semi-aquatic (Fig. 2a). The number of species with each adult diet can be found in Fig. 2b and the number of species with each larval diet can be found in Fig. 2c. Traits were assigned to the lowest taxonomic level possible and species-level traits were found for 190 species.
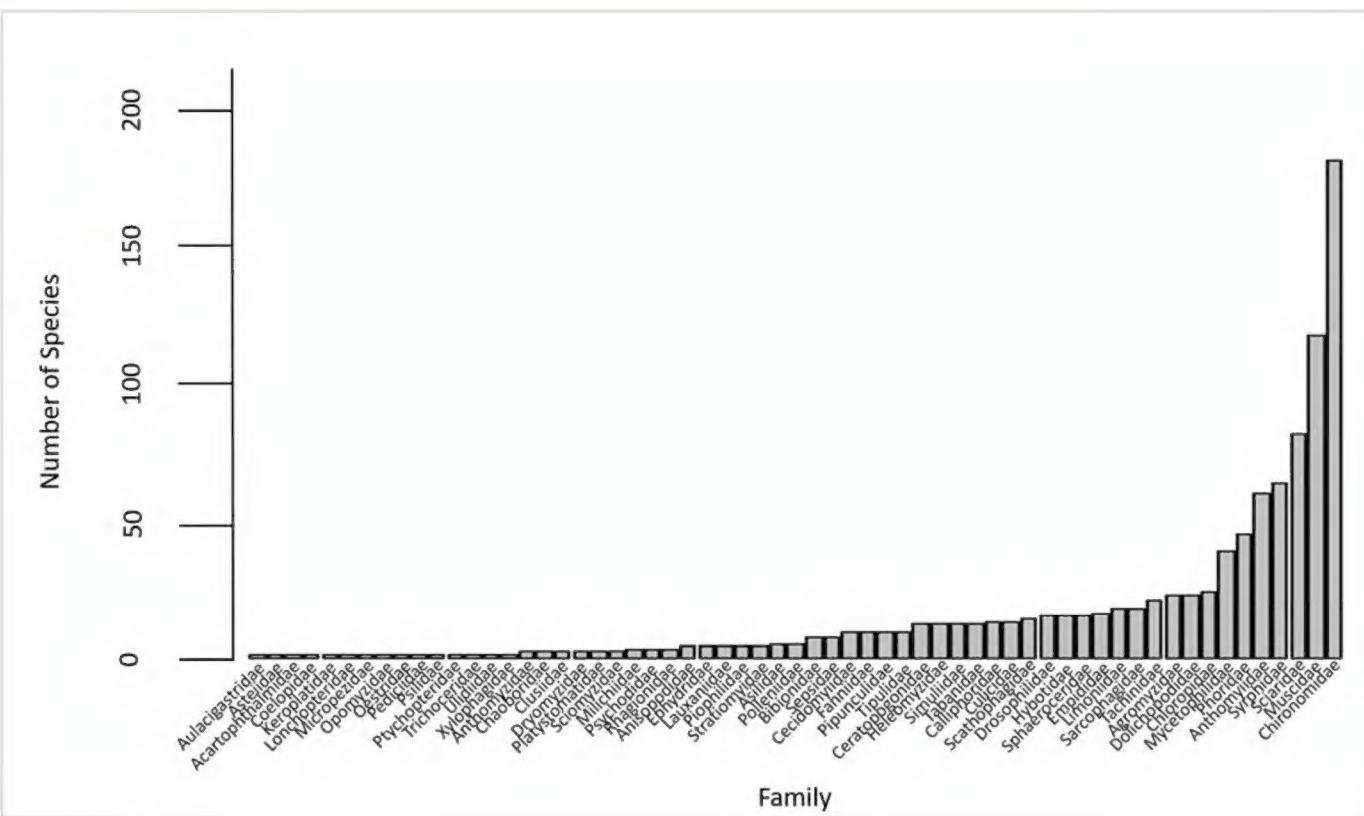
Figure 1. doi

The number of Diptera species from Canada and Greenland of each family that is present in the dataset. The selected species included here also have multiple occurrence records and DNA barcode sequences publicly available on BOLD. For several families, only one species is included. Chironomidae is the largest family within the dataset, with 182 species present.
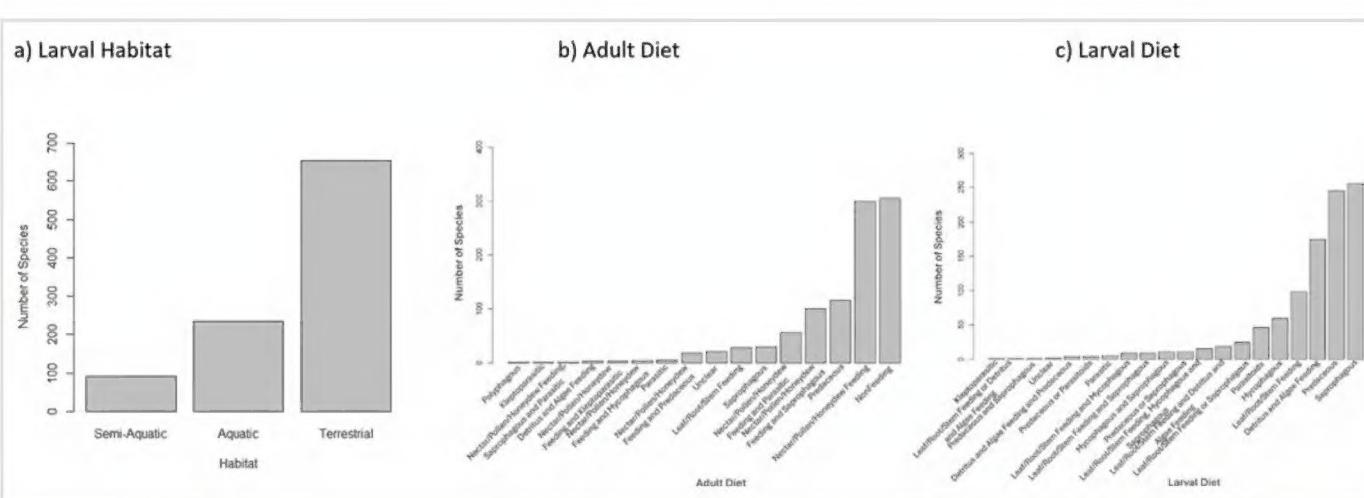
Figure 2. doi

The number of Diptera species from Canada and Greenland in this dataset having each trait category for: **a** larval habitat; **b** adult diet; **c** larval diet. The majority of species included in the dataset are terrestrial. Most species are non-feeding at the adult life stage and few are polyphagous or kleptoparasitic as adults. The majority of species included in the dataset are saprophagous as larva and few are kleptoparasitic.

Temporal coverage

Notes: The literature searches were conducted from April 2021 to April 2024. The papers referenced were published from 1889 to 2023.

Usage licence

Usage licence: Creative Commons Public Domain Waiver (CC-Zero)

Data resources

Data package title: CanFlyet: Habitat Zone and Diet Trait Dataset for Diptera Species of Canada and Greenland

Number of data sets: 1

Data set name: CanFlyet: Habitat Zone and Diet Trait Dataset for Diptera Species of Canada and Greenland

Download URL: https://datadryad.org/stash/dataset/doi:10.5061/dryad.fqz612jwx

Data format: csv, xlsx, tsv

Data format version: xlsx is version 2450. Csv and tsv are UTF-8.

Description: The dataset presents larval habitat, adult diet and larval diet for 981 species across 380 genera, 34 subfamilies and 61 families. A table showing the column labels and descriptions is provided in below. The Diptera species included are from Canada and Greenland and have multiple occurrence records and DNA sequence data on BOLD. The traits were assigned to taxa using a series of literature searches conducted between April 2021 and April 2024. This dataset can be used by researchers to determine biological traits for Diptera species and conduct research that involves ecological trait data.

Column label	Column description
id	The serial number of the record.
taxonRank	The taxonomic level the record has been identified to.
order	The order-level taxonomic classification.
family	The family-level taxonomic classification.
subfamily	The subfamily-level taxonomic classification.
genus	The genus-level taxonomic classification.
specificEpithet	The species-level taxonomic classification.
scientificName	The scientific name of the species.

habitat	The larval habitat in which the taxon is found. Taxa can be assigned terrestrial (taxa that live primarily in land habitats), aquatic (taxa that live primarily in waterbodies or associated habitats) or semi-aquatic (taxa that primarily live in wet habitats and require high levels of moisture and some elements of both terrestrial and aquatic habitats).
traitName	The name of the biological trait. The traits included are adultDiet and larvalDiet.
traitValue	The trait category assigned to the species. Taxa can be assigned predaceous (taxa that prey on insects or other animals), mycophagous (those that feed on fungi), saprophagous (those that feed on decaying organic matter), nectar/ pollen/honeydew feeding (taxa that primarily feed on nectar, pollen and/or honeydew), parasitic (those that live and feed in or on an organism of another species), parasitoid (organism whose young develop within a host, eventually resulting in host death), leaf/root/stem feeding (those that feed on the leaves, roots and/or stems of plants), detritus and algae feeding (those that feed on small particles of algae and detritus), kleptoparasitic (those that take food and resources from another species), polyphagous (those that engage in more than three of the other feeding modes), non-feeding (those that do not feed at the specified life stage) and unclear (taxa for which there is not enough information to make a trait assignment). The taxa can also be assigned a combination of these categories. In this case, the categories are separated by "and". For example, "saprophagous and predaceous". For cases where it is unclear which diet is more common, the categories are separated by "or". For example, "saprophagous or predaceous".
traitAssignmentLevelHabitat	The taxonomic level from which the habitat assignment was obtained and assigned.
traitAssignmentLevelAdultDiet	The taxonomic level from which the adult diet assignment was obtained and assigned.
traitAssignmentLevelLarvalDiet	The taxonomic level from which the larval diet assignment was obtained and assigned.
notes	Additional information on the habitat and diet of the taxa. Information is provided for the lowest taxonomic level used for the trait assignments if available.
basisOfRecord	From where the trait data were obtained.
Number_of_references_consulted	The number of sources referenced when obtaining the trait information for each taxon.
references	The title and DOI of the reference used for the trait assignment. Full references are included in the manuscript.

Additional information

For this study, we created a dataset containing larval habitat and adult and larval diet for 981 Diptera species across 61 families found in Canada and Greenland. This dataset also contains trait information for nine taxa only identified to the genus level. This dataset allows for assignment of traits to a large variety of species. This is a valuable resource for bioinformatics work, as well as for ecological studies. This dataset was originally used to determine the relationship between biological traits and population genetic structure (Majoros et al. 2023) and can be applied to a wide range of other studies and research. These traits can also be linked to other types of data, such as species occurrence or community composition data.

The species included in this dataset possessed a wide variety of different habitat and dietary requirements. Most species were terrestrial, non-feeding at the adult life stage or saprophagous at the larval stage. Other diets were far less represented, such as those with a kleptoparasitic life stage.

Trait data have been used to investigate various ecological and evolutionary patterns and topics, such as community assembly (Kraft and Ackerly 2010, Zhang et al. 2020), molecular evolution (May et al. 2020), phylogenetic assemblage structure (Barnagaud et al. 2014) and how populations change over time or in response to ecological change (De Palma et al. 2015, Coulthard et al. 2019). Similarly, this dataset can be used to answer a variety of questions and is a useful tool for future research. It can be used for studies that relate to ecology, evolution, biogeography and conservation of fly species and include the habitat and diet of flies. The dataset can be used to investigate the relationship between traits and population genetic structure or phylogenetic community structure. The dataset could also be used to find relationships between traits, such as in studies like Freire et al. (2021). Visualisations are provided in Fig. 3, Fig. 4 and Fig. 5 of the phylogenetic relationships between species possessing different traits. This approach could be built upon to look at the relationship between traits and evolutionary history, such as in studies like Rainford and Mayhew (2015). The trait dataset can also be used to investigate the diversity of different fly traits in an area of interest, to determine how these traits impact occurrence patterns, how traits may relate to responses to environmental change or to determine which species may be most likely to colonise these areas in the future. Trait data could also prove useful for biomonitoring of insect diversity in agricultural areas, such as research into which farming practices are associated with a balance of ecological functions in associated insect communities.

While this dataset provides useful trait information for a large number of Diptera species, more research is needed to expand our knowledge of this ecologically important group. This dataset can be expanded upon to include other important traits, such as adult habitat, more detail about foods comprising the larval and adult diets, reproductive traits, mobility traits and vector status. With the inclusion of additional traits, the dataset can be used to answer an even larger variety of biological questions. It is also important to consider the diversity of Diptera species and we note that, even within genera, species

may differ in habitat and diet requirements. While we made trait assignments based on previous research and taxonomy, further research should be done to explore uncertainties, add detail and achieve species-level data. Diptera habitat preferences can also vary across study areas and further research should be done to understand the ecology of Diptera species in any given geographic area. While it is hard to capture these details in a general dataset, the data provided in this study provide a valuable resource for a range of scientific studies, as well as a building block for future work.

Habitat
- Terrestrial
- Aquatic
- Semi-aquatic
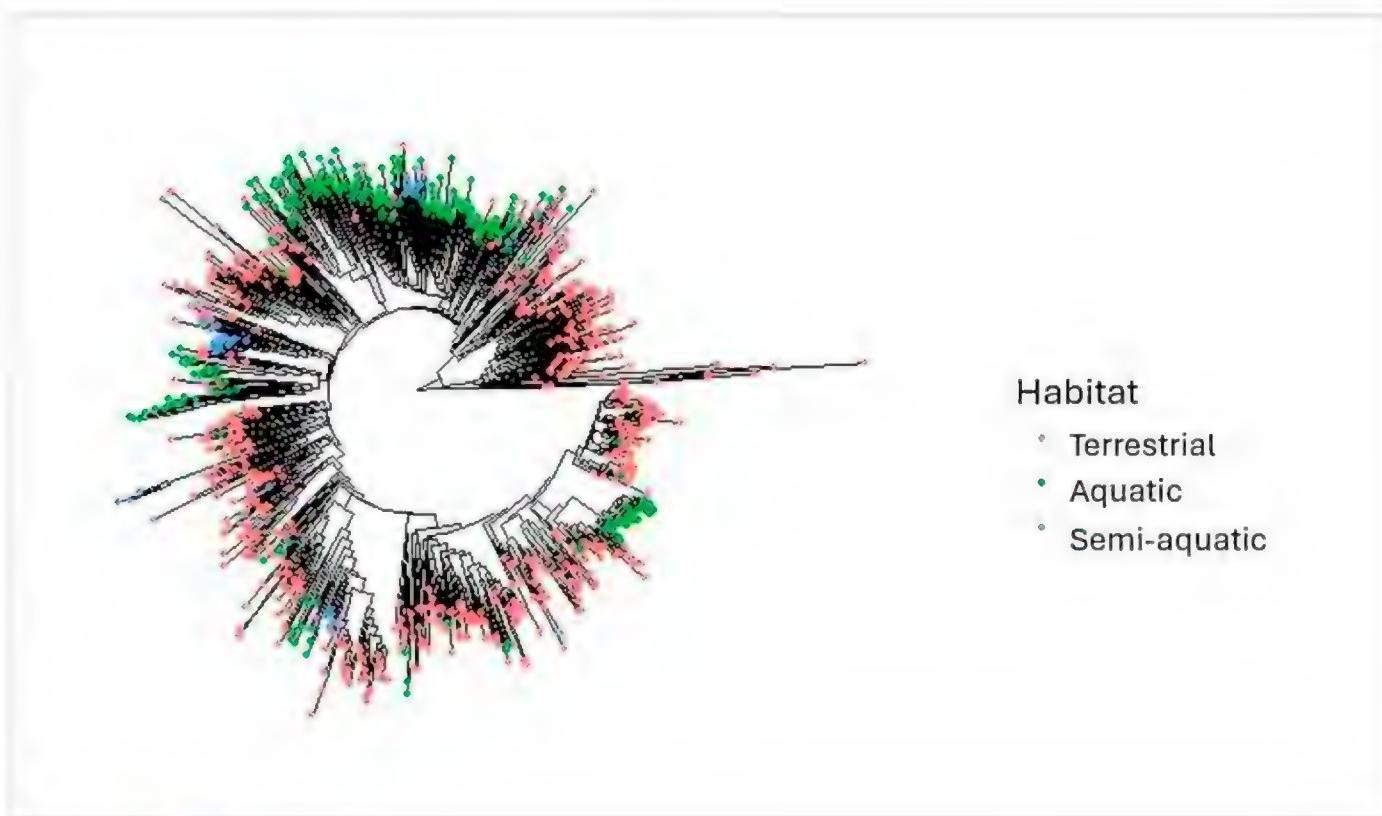
Figure 3. doi

Phylogenetic tree of Diptera species from Canada and Greenland included in this dataset; these species also have multiple occurrence records and DNA barcode sequences publicly available on BOLD. The different colours of the tips represent species with different larval habitat categories. The traits provided in the dataset could be used to investigate the relationship between traits and evolutionary history. These Maximum Likelihood trees were created using cytochrome c oxidase subunit 1 (CO1) sequences from BOLD and functions from the package phangorn version 2.11.1 (Schliep 2011) in the R programming language (R Core Team 2023). The tips were coloured using the package ggtree version 3.6.2 (Yu et al. 2017). The code for how to create this tree is provided at https://github.com/S-Majoros/Diptera_Dataset_Phylogenetic_Tree.

In summary, we share a trait dataset containing habitat and diet information for Diptera taxa from Canada and Greenland that also have multiple occurrence records and DNA barcoding sequences available on BOLD. This dataset is publicly available on Data Dryad (https://datadryad.org/stash/dataset/doi:10.5061/dryad.fqz612jwx) and is also available on GitHub (https://github.com/S-Majoros/Diptera_Dataset_Phylogenetic_Tree). The file is available in csv, xlsx and tsv formats in order to make it as accessible as possible for future users. This dataset is publicly available so that other researchers can use it to conduct further studies, answer more questions and improve our knowledge of this ecologically important group.

Adult Diet
- Nectar/Pollen/Honeydew Feeding
- Non-Feeding
- Predaceous
- Saprophagous
- Leaf/Root/Stem Feeding
- Unclear
- Parasitic
- Detritus and Algae Feeding
- Polyphagous
- Nectar/Pollen/Honeydew Feeding and Parasitic
- Nectar/Pollen/Honeydew Feeding and Predaceous
- Nectar/Pollen/Honeydew Feeding and Saprophagous
- Nectar/Pollen/Honeydew Feeding and Mycophagous
- Nectar/Pollen/Honeydew Feeding and Kleptoparasitic
- Nectar/Pollen/Honeydew Feeding, Saprophagous and Parasitic

Figure 4. doi

Phylogenetic tree of Diptera species from Canada and Greenland included in this dataset; these species also have multiple occurrence records and DNA barcode sequences publicly available on BOLD. The different colours of the tips represent species with different adult diet categories. The traits provided in the dataset could be used to investigate the relationship between traits and evolutionary history. These Maximum Likelihood trees were created using cytochrome c oxidase subunit 1 (CO1) sequences from BOLD and functions from the package phangorn version 2.11.1 (Schliep 2011) in the R programming language (R Core Team 2023). The tips were coloured using the package ggtree version 3.6.2 (Yu et al. 2017). The code for how to create this tree is provided at https://github.com/S-Majoros/Diptera_Dataset_Phylogenetic_Tree.

Larval Diet
- Saprophagous
- Detritus and Algae Feeding
- Predaceous
- Leaf/Root/Stem Feeding
- Parasitoid
- Mycophagous
- Unclear
- Parasitic
- Kleptoparasitic
- Detritus and Algae Feeding and Predaceous
- Mycophagous and Saprophagous
- Leaf/Root/Stem Feeding and Detritus and Algae Feeding
- Leaf/Root/Stem Feeding and Mycophagous
- Leaf/Root/Stem Feeding and Saprophagous
- Predaceous and Saprophagous
- Leaf/Root/Stem Feeding, Mycophagous and Saprophagous
- Predaceous or Saprophagous
- Leaf/Root/Stem Feeding or Saprophagous
- Predaceous or Parasitoid
- Leaf/Root/Stem Feeding or Detritus and Algae Feeding
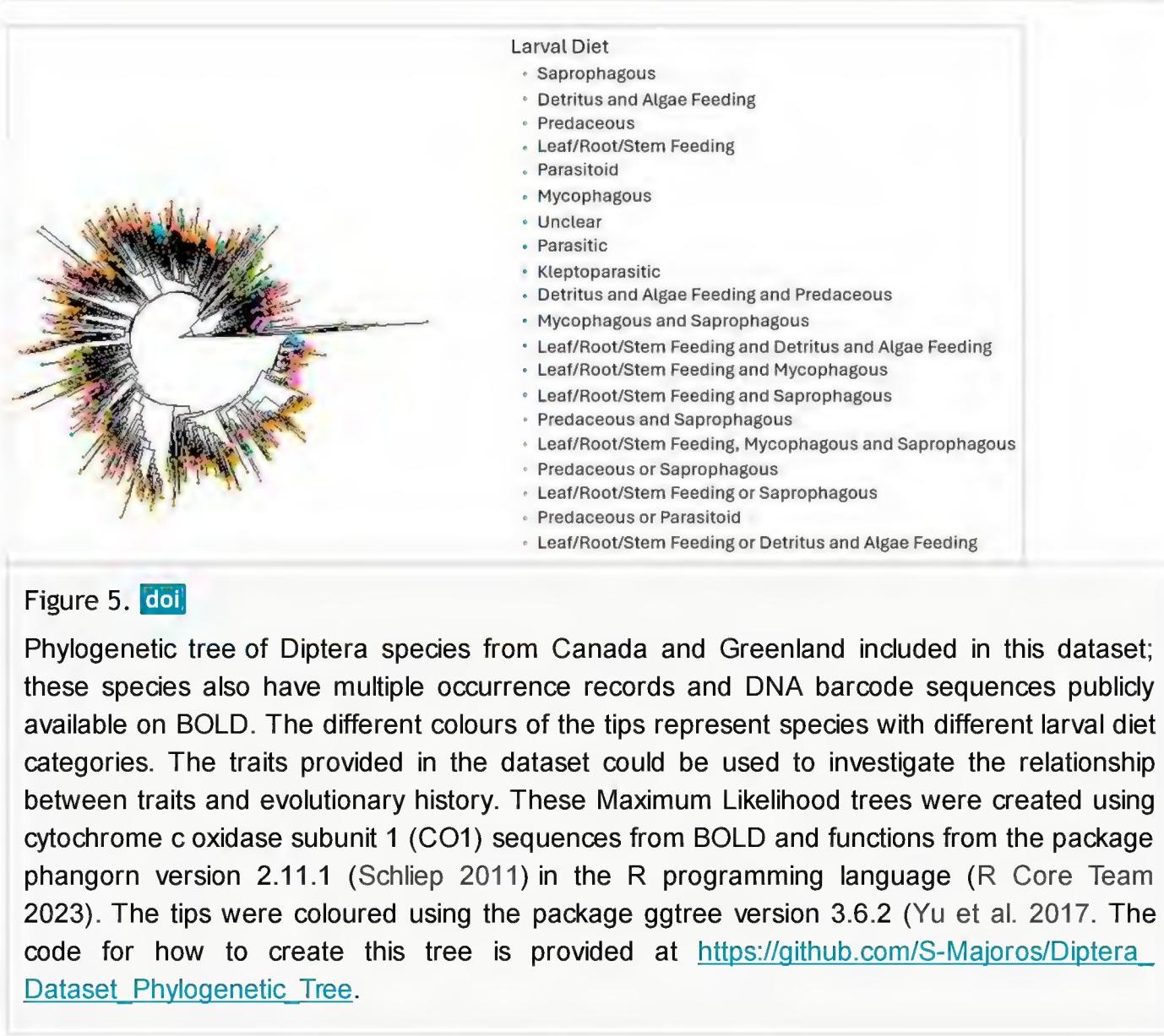
Figure 5. doi

Phylogenetic tree of Diptera species from Canada and Greenland included in this dataset; these species also have multiple occurrence records and DNA barcode sequences publicly available on BOLD. The different colours of the tips represent species with different larval diet categories. The traits provided in the dataset could be used to investigate the relationship between traits and evolutionary history. These Maximum Likelihood trees were created using cytochrome c oxidase subunit 1 (CO1) sequences from BOLD and functions from the package phangorn version 2.11.1 (Schliep 2011) in the R programming language (R Core Team 2023). The tips were coloured using the package ggtree version 3.6.2 (Yu et al. 2017. The code for how to create this tree is provided at https://github.com/S-Majoros/Diptera_Dataset_Phylogenetic_Tree.

Acknowledgements

Author contributions

Samantha E. Majoros – project design, dataset construction, manuscript writing and editing.

Tyler A. Elliott – dataset design, data validation, manuscript editing.

Sarah J. Adamowicz – project design, manuscript editing.

References

- Adler PH, Courtney GW (2019) Ecological and societal services of aquatic Diptera. Insects 10 (3): 70. https://doi.org/10.3390/insects10030070
- Baloğlu B, Clews E, Meier R (2018) NGS barcoding reveals high resistance of a hyperdiverse chironomid (Diptera) swamp fauna against invasion from adjacent freshwater reservoirs. Frontiers in Zoology 15 (1): 31. https://doi.org/10.1186/s12983-018-0276-7
- Barnagaud J, Kissling WD, Sandel B, Eiserhardt WL, Şekercioğlu CH, Enquist BJ, Tsirogiannis C, Svenning J (2014) Ecological traits influence the phylogenetic structure of bird species co-occurrences worldwide. Ecology Letters 17 (7): 811-820. https://doi.org/10.1111/ele.12285
- Bedini S, Flamini G, Cosci F, Ascrizzi R, Echeverria MC, Guidi L, Landi M, Lucchi A, Conti B (2017) *Artemisia* spp. essential oils against the disease-carrying blowfly *Calliphora vomitoria*. Parasites & Vectors 10 (1): 80. https://doi.org/10.1186/s13071-017-2006-y
- Bragard C, Dehnen-Schmutz K, Di Serio F, Gonthier P, Jacques M, Miret JA, Justesen AF, Magnusson CS, Milonas P, Navas-Cortes JA, Parnell S, Potting R, Reignault PL, Thulke H, Van der Werf W, Civera AV, Yuen J, Zappalà L, Czwienczek E, Streissl F, MacLeod A (2020) Pest categorisation of *Liriomyza bryoniae*. EFSA Journal 18 (3). https://doi.org/10.2903/j.efsa.2020.6038

- Coulthard E, Norrey J, Shortall C, Harris WE (2019) Ecological traits predict population changes in moths. Biological Conservation 233: 213-219. https://doi.org/10.1016/j.biocon.2019.02.023
- Darwin Core Maintenance Group (2023) Darwin Core List of Terms. Biodiversity Information Standards (TDWG). http://rs.tdwg.org/dwc/doc/list/2023-09-18
- De Palma A, Kuhlmann M, Roberts SP, Potts SG, Börger L, Hudson LN, Lysenko I, Newbold T, Purvis A (2015) Ecological traits affect the sensitivity of bees to land-use pressures in European agricultural landscapes. Journal of Applied Ecology 52 (6): 1567-1577. https://doi.org/10.1111/1365-2664.12524
- deWaard JR, Levesque-Beaudin V, deWaard SL, Ivanova NV, McKeown JT, Miskie R, Naik S, Perez KH, Ratnasingham S, Sobel CN, Sones JE, Steinke C, Telfer AC, Young AD, Young MR, Zakharov EV, Hebert PD (2018) Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. Genome 62 (3): 85-95. https://doi.org/10.1139/gen-2018-0093
- Freire J,G, Silva T, Oliveira H, Collier C, Rodrigues HP, Dias JP, Santos JP, Marini-Filho OJ, Freitas AV, Smilanich AM, Dyer LA, Diniz IR (2021) Good things come in larger packages: Size matters for adult fruit-feeding butterfly dispersal and larval diet breadth. Diversity 13 (12): 664. https://doi.org/10.3390/d13120664
- GBIF.org (2024) GBIF Home Page. https://www.gbif.org. Accessed on: 2024-5-28.
- Geiger MF, Moriniere J, Hausmann A, Haszprunar G, Wägele W, Hebert PD, Rulik B (2016) Testing the Global Malaise Trap Pogram-How well does the current barcode reference library identify flying insects in Germany? Biodiversity Data Journal 4 (4). https://doi.org/10.3897/BDJ.4.e10671
- Global Malaise Trap Program (2023) Global Malaise Trap Program. http://globalmalaise.org/homecopy/
- Gossner MM, Simons NK, Achtziger R, Blick T, Dorow WH, Dziock F, Köhler F, Rabitsch W, Weisser WW (2015) A summary of eight traits of Coleoptera, Hemiptera, Orthoptera and Araneae, occurring in grasslands in Germany. Scientific Data 2 (1). https://doi.org/10.1038/sdata.2015.13
- Hagge J, Müller J, Birkemoe T, Buse J, Christensen RH, Gossner MM, Gruppe A, Heibl C, Jarzabek-Müller A, Seibold S, Siitonen J, Soutinho JG, Sverdrup-Thygeson A, Thorn S, Drag L (2021) What does a threatened saproxylic beetle look like? Modelling extinction risk using a new morphological trait database. Journal of Animal Ecology 90 (8): 1934-1947. https://doi.org/10.1111/1365-2656.13512
- Hebert PD, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270 (1512): 313-321. https://doi.org/10.1098/rspb.2002.2218
- Herberstein ME, McLean DJ, Lowe E, Wolff JO, Khan MK, Smith K, Allen AP, Bulbert M, Buzatto BA, Eldridge MD, Falster D, Winzer LF, Griffith SC, Madin JS, Narendra A, Westoby M, Whiting MJ, Wright IJ, Carthey AJ (2022) AnimalTraits – a curated animal trait database for body mass, metabolic rate and brain size. Scientific Data 9 (1): 265. https://doi.org/10.1038/s41597-022-01364-9
- Hörren T, Sorg M, Hallmann CA, Zizka VM, Ssymank A, Noll NW, Schäffler L, Scherber C (2022) A universal insect trait tool (ITT, v1.0) for statistical analysis and evaluation of biodiversity research data. bioRxiv https://doi.org/10.1101/2022.01.25.477751

- Kraft NJ, Ackerly DD (2010) Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. Ecological Monographs 80 (3): 401-422. https://doi.org/10.1890/09-1672.1
- Lin X, Stur E, Ekrem T (2015) Exploring genetic divergence in a species-rich insect genus using 2790 DNA barcodes. *PLoS One* 10 (9). https://doi.org/10.1371/journal.pone. 0138993
- Majoros SE, Adamowicz SJ, Cottenie K (2023) Novel pipeline for large-scale comparative population genetics. bioRxiv https://doi.org/10.1101/2023.01.23.524574
- Marshall SA (2012) Flies. The natural history and diversity of Diptera. Firefly Books Ltd.
- May JA, Feng Z, Orton MG, Adamowicz SJ (2020) The effects of ecological traits on the rate of molecular evolution in ray-finned fishes: a multivariable approach. Journal of Molecular Evolution 88 (8-9): 689-702. https://doi.org/10.1007/s00239-020-09967-9
- Mesmin X, Vincent M, Tricault Y, Estorgues V, Daniel L, Cortesero A, Faloya V, Le Ralec A (2019) Assessing the relationship between pest density and plant damage: a case study with the belowground herbivore *Delia radicum* (Diptera: Anthomyiidae) on broccoli. Applied Entomology and Zoology 54 (2): 155-165. https://doi.org/10.1007/ s13355-019-00607-3
- Middleton-Welling J, Dapporto L, García-Barros E, Wiemers M, Nowicki P, Plazio E, Bonelli S, Zaccagno M, Šašić M, Liparova J, Schweiger O, Harpke A, Musche M, Settele J, Schmucki R, Shreeve T (2020) A new comprehensive trait database of European and Maghreb butterflies, Papilionoidea. Scientific Data 7 (1): 351. https://doi.org/10.1038/ s41597-020-00697-7
- Molinatto G, Demichelis S, Bodino N, Giorgini M, Mori N, Bosco D (2020) Biology and prevalence in Northern Italy of *Verrallia aucta* (Diptera, Pipunculidae), a parasitoid of *Philaenus spumarius* (Hemiptera, Aphrophoridae), the main vector of *Xylella fastidiosa* in Europe. Insects 11 (9): 607. https://doi.org/10.3390/insects11090607
- Olafson PU, Aksoy S, Attardo GM, Buckmeier G, Chen X, Coates CJ, Davis M, Dykema J, Emrich SJ, Friedman M, Holmes CJ, Ioannidis P, Jansen EN, Jennings EC, Lawson D, Martinson EO, Maslen GL, Meisel RP, Murphy TD, Nayduch D, Nelson DR, Oyen KJ, Raszick TJ, Ribeiro JM, Robertson HM, Rosendale AJ, Sackton TB, Saelao P, Swiger SL, Sze S, Tarone AM, Taylor DB, Warren WC, Waterhouse RM, Weirauch MT, Werren JH, Wilson RK, Zdobnov EM, Benoit JB (2021) The genome of the stable fly, *Stomoxys calcitrans*, reveals potential mechanisms underlying reproduction, host interactions, and novel targets for pest control. BMC Biology 19 (1): 41. https://doi.org/10.1186/ s12915-021-00975-9
- Parr CL, Dunn RR, Sanders NJ, Weiser MD, Photakis M, Bishop TR, Fitzpatrick MC, Arnan X, Baccaro F, Brandão CR, Chick L, Donoso DA, Fayle TM, Gómez C, Grossman B, Munyai TC, Pacheco R, Retana J, Robinson A, Sagata K, Silva RR, Tista M, Vasconcelos H, Yates M, Gibb H (2016) GlobalAnts: a new database on the geography of ant traits (Hymenoptera: Formicidae). Insect Conservation and Diversity 10 (1): 5-20. https://doi.org/10.1111/icad.12211
- Rainford JL, Mayhew PJ (2015) Diet evolution and clade richness in Hexapoda: a phylogenetic study of higher taxa. The American Naturalist 186 (6): 771-791. https:// doi.org/10.1086/683461
- Ratnasingham S, Hebert PD (2007) Bold: the barcode of life data system. Molecular Ecology Notes 7 (3): 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678

- Ratnasingham S, Hebert PD (2013) A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. PLOS ONE 8 (7).
- R Core Team (2023) R: A language and environment for statistical computing. http://www.R-project.org/
- Sarremejane R, Cid N, Stubbington R, Datry T, Alp M, Cañedo-Argüelles M, Cordero-Rivera A, Csabai Z, Gutiérrez-Cánovas C, Heino J, Forcellini M, Millán A, Paillex A, Pařil P, Polášek M, Tierno de Figueroa JM, Usseglio-Polatera P, Zamora-Muñoz C, Bonada N (2020) DISPERSE, a trait database to assess the dispersal potential of European aquatic macroinvertebrates. Scientific Data 7 (1): 386. https://doi.org/10.1038/s41597-020-00732-7
- Schliep KP (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27 (4): 592-593. https://doi.org/10.1093/bioinformatics/btq706
- Shamshad A (2010) The development of integrated pest management for the control of mushroom sciarid flies, *Lycoriella ingenua* (Dufour) and *Bradysia ocellaris* (Comstock), in cultivated mushrooms. Pest Management Science 66 (10): 1063-1074. https://doi.org/10.1002/ps.1987
- Shirey V, Larsen E, Doherty A, Kim CA, Al-Sulaiman FT, Hinolan JD, Itliong MG, Naive MA, Ku M, Belitz M, Jeschke G, Barve V, Lamas G, Kawahara AY, Guralnick R, Pierce NE, Lohman DJ, Ries L (2022) LepTraits 1.0 A globally comprehensive dataset of butterfly traits. Scientific Data 9 (1): 382. https://doi.org/10.1038/s41597-022-01473-5
- Tiusanen M, Hebert PD, Schmidt NM, Roslin T (2016) One fly to rule them all – muscid flies are the key pollinators in the Arctic. Proceedings of the Royal Society: Biological Sciences 283 (1839). https://doi.org/10.1098/rspb.2016.1271
- Twardochleb L, Hiltner E, Pyne M, Zarnetske P (2021) Freshwater insects CONUS: A database of freshwater insect occurrences and traits for the contiguous United States. Global Ecology and Biogeography 30 (4): 826-841. https://doi.org/10.1111/geb.13257
- United States Environmental Protection Agency (2012) Freshwater Biological Trait Database. (Report No. EPA/600/R-11/038F). https://cfpub.epa.gov/ncea/global/recordisplay.cfm?deid=241813
- van der Plas F, van Klink R, Manning P, Olff H, Fischer M (2017) Sensitivity of functional diversity metrics to sampling intensity. Methods in Ecology and Evolution 8 (9): 1072-1080. https://doi.org/10.1111/2041-210X.12728
- Vieira NK, Poff NL, Carlisle DM, Moulton II SR, Koski ML, Kondratieff BC (2006) A database of lotic invertebrate traits for North America. United States Geographical Survey URL: https://pubs.usgs.gov/ds/ds187/pdf/ds187.pdf
- Waller JT, Willink B, Tschol M, Svensson EI (2019) The odonate phenotypic database, a new open data resource for comparative studies of an old insect order. Scientific Data 6 (1): 316. https://doi.org/10.1038/s41597-019-0318-9
- Wearing CH, Marshall RR, Attfield B, Colhoun C (2013) Phenology and distribution of the apple leafcurling midge (*Dasineura mali* (Kieffer)) (Diptera: Cecidomyiidae) and its natural enemies on apples under biological and integrated pest management in Central Otago, New Zealand. New Zealand Entomologist 36 (2): 87-106. https://doi.org/10.1080/00779962.2012.712887
- Yu G, Smith D, Zhu H, Guan Y, Lam TT (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8 (1): 28-36. https://doi.org/10.1111/2041-210X.12628

- Zhang J, Swenson NG, Liu J, Liu M, Qiao X, Jiang M (2020) A phylogenetic and trait-based analysis of community assembly in a subtropical forest in central China. Ecology and Evolution 10 (15): 8091-8104. https://doi.org/10.1002/ece3.6465

Supplementary material

Suppl. material 1: References for Diptera habitat and diet traits `doi`

Authors: Samantha E. Majoros, and Sarah J. Adamowicz
Data type: Literature References
Brief description: This file contains the references used to determine habitat and diet for Diptera species from Canada and Greenland.
Download file (652.00 kb)

Endnotes

*1   https://figshare.com/articles/dataset/
      DISPERSE_a_trait_database_to_assess_the_dispersal_potential_of_European_a
      quatic_macroinvertebrates/12417251/1
*2   https://cfpub.epa.gov/ncea/global/traits/search.cfm
*3   https://animaltraits.org/
*4   https://portal.edirepository.org/nis/mapbrowse?packageid=edi.481.5
*5   https://pubs.er.usgs.gov/publication/ds187
*6   https://butterflytraits.github.io/European-Butterfly-Traits/index.html
*7   http://globalants.org/
*8   http://www.odonatephenotypicdatabase.org/shiny/odonates/?
      inputs&choose_species=%22%22
*9   https://datadryad.org/stash/dataset/doi:10.5061/dryad.2fqz612p3
*10   https://datadryad.org/stash/dataset/doi:10.5061/dryad.1fn46
*11   https://datadryad.org/stash/dataset/doi:10.5061/dryad.53ds2
*12   https://springernature.figshare.com/collections/
      LepTraits_1_0_A_globally_comprehensive_dataset_of_butterfly_traits/5899187/1
*13   https://www.biorxiv.org/content/10.1101/2022.01.25.477751v1.full